

## University of Groningen

### LOFAR Lightning Imaging

Hare, B. M.; Scholten, O.; Bonardi, A.; Buitink, S.; Corstanje, A.; Ebert, U.; Falcke, H.; Horandel, J. R.; Leijnse, H.; Mitra, P.

*Published in:*  
Journal of geophysical research-Atmospheres

*DOI:*  
[10.1002/2017JD028132](https://doi.org/10.1002/2017JD028132)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2018

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Hare, B. M., Scholten, O., Bonardi, A., Buitink, S., Corstanje, A., Ebert, U., Falcke, H., Horandel, J. R., Leijnse, H., Mitra, P., Mulrey, K., Nelles, A., Rachen, J. P., Rossetto, L., Rutjes, C., Schellart, P., Thoudam, S., Trinh, T. N. G., ter Veen, S., & Winchen, T. (2018). LOFAR Lightning Imaging: Mapping Lightning With Nanosecond Precision. *Journal of geophysical research-Atmospheres*, 123(5), 2861-2876.  
<https://doi.org/10.1002/2017JD028132>

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*



## RESEARCH ARTICLE

10.1002/2017JD028132

## Key Points:

- We can map lightning leader evolution with the LOFAR radio telescope in three dimensions
- The position accuracy can be on the order of meters above a surface area of 3,200 km<sup>2</sup> and the timing accuracy is around 5 ns
- LOFAR is sensitive to the current distribution in space within individual leader steps

## Correspondence to:

B. M. Hare,  
B.H.Hare@rug.nl

## Citation:

Hare, B. M., Scholten, O., Bonardi, A., Buitink, S., Corstanje, A., Ebert, U., et al. (2018). LOFAR lightning imaging: Mapping lightning with nanosecond precision. *Journal of Geophysical Research: Atmospheres*, 123, 2861–2876. <https://doi.org/10.1002/2017JD028132>

Received 30 NOV 2017

Accepted 12 FEB 2018

Accepted article online 16 FEB 2018

Published online 12 MAR 2018

## LOFAR Lightning Imaging: Mapping Lightning With Nanosecond Precision

B. M. Hare<sup>1</sup>, O. Scholten<sup>1,2</sup>, A. Bonardi<sup>3</sup>, S. Buitink<sup>4</sup>, A. Corstanje<sup>3</sup>, U. Ebert<sup>5,6</sup>, H. Falcke<sup>2,7,8,9</sup>, J. R. Hörandel<sup>3,7</sup>, H. Leijnse<sup>10</sup>, P. Mitra<sup>4</sup>, K. Mulrey<sup>4</sup>, A. Nelles<sup>3,11</sup>, J. P. Rachen<sup>3</sup>, L. Rossetto<sup>3</sup>, C. Rutjes<sup>5</sup>, P. Schellart<sup>3,12</sup>, S. Thoudam<sup>13</sup>, T. N. G. Trinh<sup>1</sup>, S. ter Veen<sup>3,8</sup>, and T. Winchen<sup>4</sup>
<sup>1</sup>KVI-Center for Advanced Radiation Technology, University of Groningen, Groningen, Netherlands,

<sup>2</sup>Interuniversity Institute for High-Energy, Vrije Universiteit Brussel, Brussels, Belgium, <sup>3</sup>Department of Astrophysics/IMAPP, Radboud University Nijmegen, Nijmegen, Netherlands, <sup>4</sup>Astrophysical Institute, Vrije Universiteit Brussel, Brussels, Belgium, <sup>5</sup>Center for Mathematics and Computer Science, Amsterdam, Netherlands, <sup>6</sup>Department of Applied Physics, Eindhoven University of Technology, Eindhoven, Netherlands, <sup>7</sup>NIKHEF, Science Park Amsterdam, Amsterdam, Netherlands, <sup>8</sup>Netherlands Institute of Radio Astronomy (ASTRON), Dwingeloo, Netherlands, <sup>9</sup>Max-Planck-Institut für Radioastronomie, Bonn, Germany, <sup>10</sup>Royal Netherlands Meteorological Institute, De Bilt, Netherlands, <sup>11</sup>Department of Physics and Astronomy, University of California Irvine, Irvine, CA, USA, <sup>12</sup>Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA, <sup>13</sup>Department of Physics and Electrical Engineering, Linnuniversitetet, Växjö, Sweden

**Abstract** Lightning mapping technology has proven instrumental in understanding lightning. In this work we present a pipeline that can use lightning observed by the LOw-Frequency ARray (LOFAR) radio telescope to construct a 3-D map of the flash. We show that LOFAR has unparalleled precision, on the order of meters, even for lightning flashes that are over 20 km outside the area enclosed by LOFAR antennas (~3,200 km<sup>2</sup>), and can potentially locate over 10,000 sources per lightning flash. We also show that LOFAR is the first lightning mapping system that is sensitive to the spatial structure of the electrical current during individual lightning leader steps.

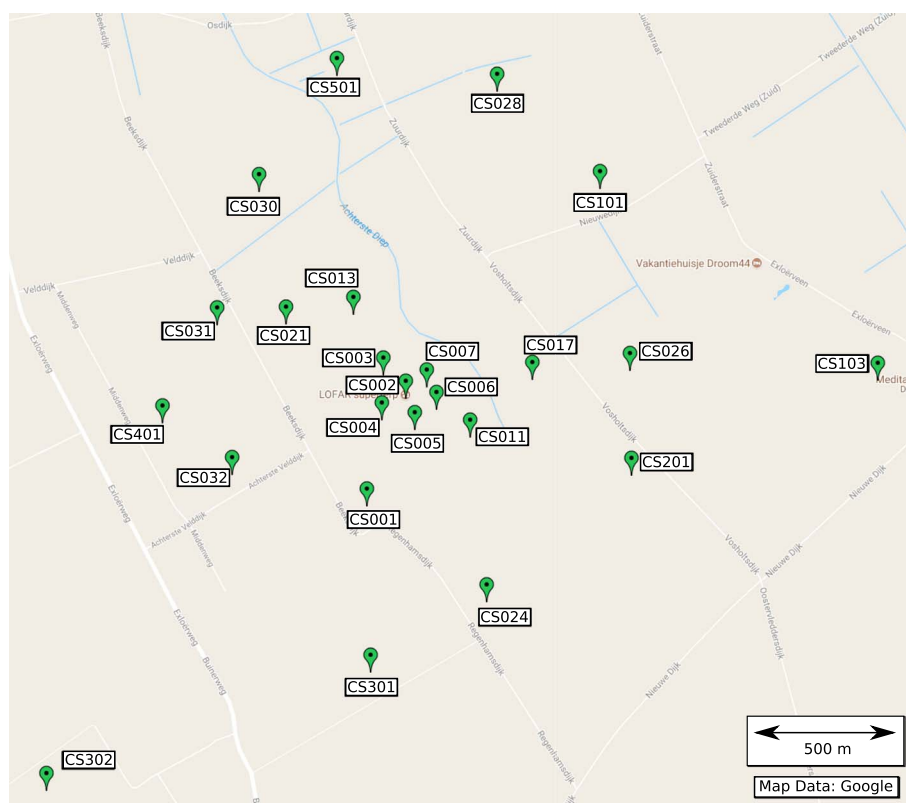
## 1. Introduction

Lightning mapping technology has improved rapidly in recent years and has become indispensable to lightning research. See Edens et al. (2012), Pilkey et al. (2014), and Rison et al. (2016) for a few examples. Most lightning mapping systems fall under two general categories: lightning mapping arrays (LMAs) and interferometers. LMAs generally consist of 6 to 20 antennas with baselines of kilometers. Each antenna has a computer that records the time and power of the strongest pulse in 10 μs bins. After a lightning flash, the times of radio frequency pulses measured by each antenna can be analyzed with a time of arrival (TOA) technique, and the lightning flash can be mapped out in three spatial dimensions and time (Rison et al., 1999). Lightning interferometers consist of a small number of antennas (typically 3 or 4) separated by a small distance, around 10–20 m. The time delays between radio frequency waveforms recorded on each antenna are extracted with cross correlation. These time delays are then used to find the azimuth and zenith angles pointing back to the lightning flash (Stock et al., 2014). While lightning interferometers can only map lightning flashes in two spatial dimensions, they can locate many more sources than a LMA. For example, Stock et al. (2014) presented a 600 ms duration lightning flash that was mapped by a digital interferometer and the Langmuir LMA. The interferometer was able to locate 62,000 sources, and the LMA was only able to locate 1,100.

We have developed a pipeline that can use data measured by the LOw-Frequency ARray (LOFAR) radio telescope to map lightning (Scholten et al., 2017; van Haarlem et al., 2012). We will show that this system is capable of mapping lightning in all three spatial dimensions and time, with a relative location accuracy on the order of meters, and can potentially locate as many sources as a lightning interferometer. In section 2 we introduce the LOFAR telescope. In section 3 we present the data that we have tested our pipeline on. In section 4 we explain the pipeline we use to map lightning with LOFAR. In section 5 we show the results of mapping the lightning flash with LOFAR. In section 6 we discuss a simple error analysis of these data. Finally, we conclude in section 7.

©2018. The Authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

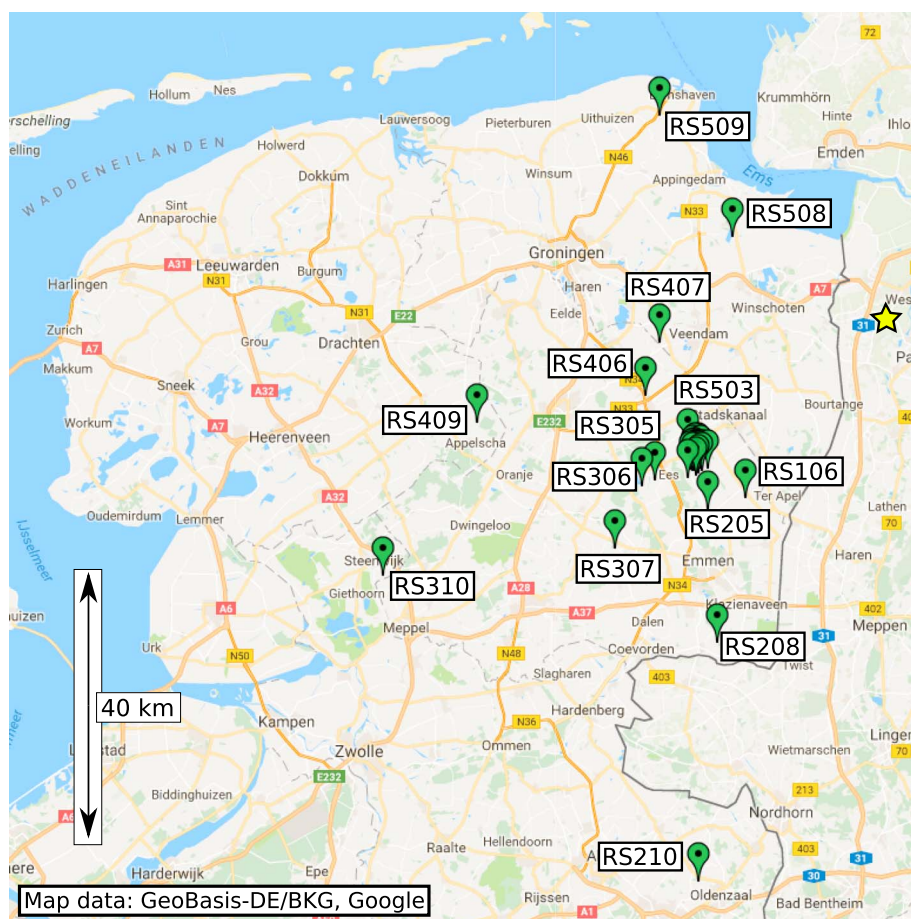


**Figure 1.** Map of the LOFAR core, based on Google maps with the names of the core stations added.

## 2. The LOFAR Radio Telescope

LOFAR is presently the world's largest radio telescope. It utilizes a phased-array design consisting of a large number of small antennas spread over a large area, designed to observe cosmic ray air showers and astronomical sources. It is a pathfinder for the Square Kilometre Array. The main core of LOFAR is in the north of the Netherlands, but there are also LOFAR stations in Germany, France, UK, Poland, Ireland, and Sweden. Figures 1 and 2 show a map of the Dutch LOFAR stations, which are split into core stations and remote stations. The Dutch LOFAR stations enclose an area of about 3,200 km<sup>2</sup>. All Dutch LOFAR stations contain 96 low-band antennas (LBAs) and 24 high-band antenna (HBA) tiles. The HBAs are not used in this work. Figure 3 shows a picture of the center of LOFAR, called the superterp. The black tiles are HBAs and the small square patches are LBAs. Notice that the LBAs are distributed inside of circles that have diameters of about 60 m. The low-band antennas are inverted V-shaped dipole antennas. They come in pairs that are collocated and orthogonal to each other so that the LBAs are sensitive to two orthogonal polarizations. One of the polarizations is oriented northwest-southeast (NW-SE), and the other is oriented northeast-southwest (NE-SW). The LBAs have a frequency range between 10 and 90 MHz, with a resonance peak at 58 MHz, and are sampled at 200 MHz. Figure 4 shows a pair of LBA dipoles and a spectrum measured by one of the LBAs. The spikes shown in the spectrum are due to local radio stations. Apart from human radio frequency interference (RFI), the noise of the LBAs is dominated by the galactic background. The raw time series data from the LBAs are saved to a transient buffer board, which can save up to 5 s of data from 48 antennas. When the transient buffer boards receive a dump command they stop recording and read the data out over a wide area network to a data processing cluster.

All the core stations are on a single GPS clock. The signal from the central clock is transmitted to each of the core stations over fiber optic cabling; however, the timing calibration of the fiber optic cables is only accurate to about 10 ns. Each of the remote stations are on separate GPS clocks and so can have relative timing offsets around hundreds of ns. Because of these additional unknown timing offsets, in section 4.2 we will use the lightning data collected by LOFAR to improve the timing calibration of the LOFAR stations. Since the timing offsets of the core stations are determined by the fiber optic cabling, the timing offsets of the core stations



**Figure 2.** Map of the Dutch LOFAR stations, based on Google maps with the names of the remote stations added. Yellow star shows the location of the lightning flash.

will change only when repair work is done on the telescope. The timing of the remote stations, however, will change continuously and will need to be recalibrated for every new set of data.

While there are 96 LBAs per station, in order to keep the amount of data manageable we use only six pairs of LBAs per station and we presently use data from only the Dutch LOFAR stations. Each station takes 30 s to save 1 s of data from each dipole. Thus, in this mode, LOFAR takes 30 min to save out 5 s of data.

### 3. Data

The data used in this work are from a single lightning flash, measured by the LOFAR radio telescope on the 12 of July 2016, at 17:34:55.100 UTC. This lightning flash was found by collecting large amounts of data during times when thunderstorms were near the area enclosed by Dutch LOFAR stations and selecting the traces that actually contained lightning data. Twelve core stations and eleven remote stations were used in this measurement. Each station had six pairs of LBA dipoles active, and 5 s of data were saved from each antenna. As we will show in section 5, the lightning flash occurred about 22 km outside the area enclosed by the active stations, 40 km north-east of the core of LOFAR and 30 km from the closest station, RS508. This location is shown in Figure 2. The flash had a duration of about 200 ms. Remote stations: RS106, RS205, RS208, RS305, RS306, RS307, RS406, RS407, RS503, RS508, and RS509 participated in recording this event.

### 4. Methodology

Because of the large amount of data, large variation in antenna baselines, and unknown timing offsets between stations, lightning mapping with LOFAR presents a significant algorithmic challenge compared to

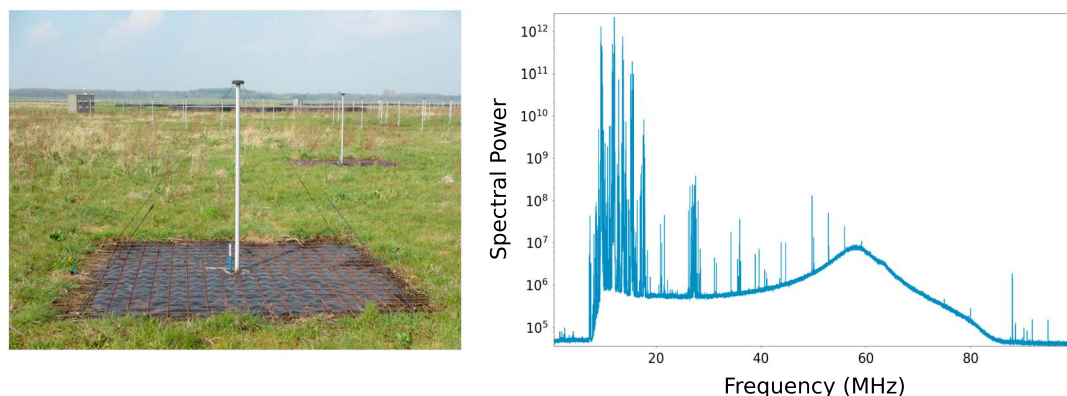




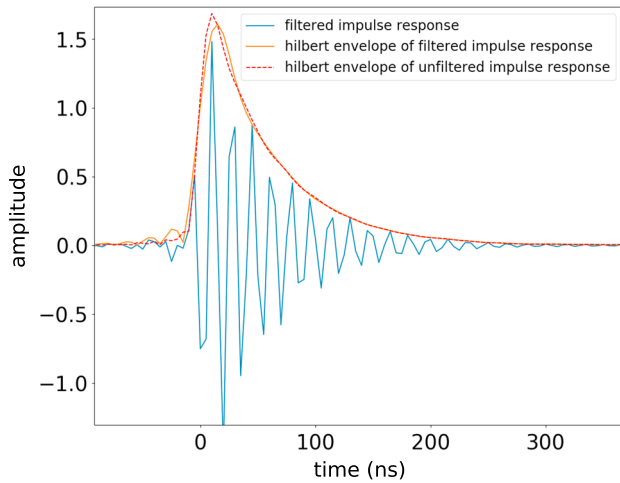
**Figure 3.** A photograph of the center of LOFAR, called the superterp, which has a diameter of about 370 m. The superterp contains the CS002, CS003, CS004, CS005, CS006, and CS007 stations. The black tiles cover the high-band antennas, and the small gray squares are the low-band antennas. Photo courtesy of ASTRON: <https://www.astron.nl/about-astron/press-public/pictures/pictures>. ©Top-Foto, Assen

traditional LMAs and interferometers. The analysis is split into three sections: pulse finding, station timing calibration, and TOA lightning mapping.

Note that our pulse-finding algorithm, described in section 4.1.2, can select traces with multiple maxima as individual pulses. For this reason, in this work we define a “pulse” as the result of our pulse-finding algorithm applied to a single antenna. The individual maxima within a pulse will be referred to as “peaks.” The physical process that produces pulses on different antennas will be called an “event.” Examples of pulses found by our algorithm are shown in Figures 7 and 8, both of which show two pulses detected on two antennas (one pulse per antenna), from the same event. Figure 8 shows data from a complex event where the two pulses have



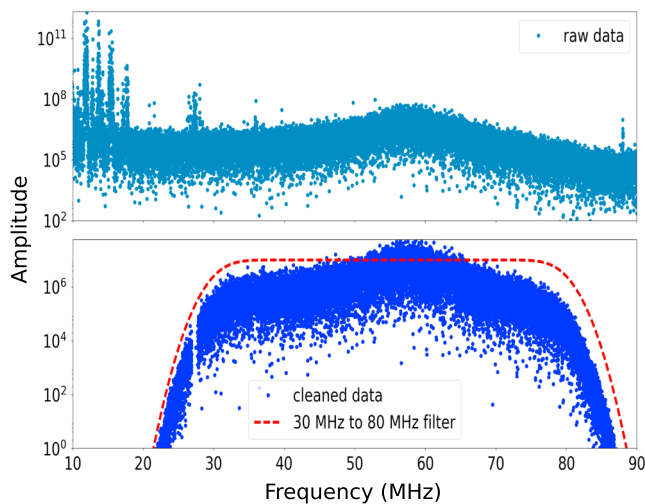
**Figure 4.** Left: A photograph of a LOFAR low-band antenna. The wire mesh on the ground is the grounding plane, and the inverted V-shaped wires is one dipole. The two dipoles are oriented toward the northwest-southeast (NW-SE) and the northeast-southwest (NE-SW) directions. Right: A typical spectrum measured by a LBA, averaged over a large number of blocks of data. The resonance peak is clearly seen at 58 MHz, as well as radio frequency interference mostly below 30 MHz.



**Figure 5.** The oscillating trace data show the real component of the impulse response after filtering, in arbitrary units. The solid line shows the absolute value of the impulse response, which is the Hilbert envelope, after filtering. The dashed line shows the absolute value of the impulse response before filtering. The difference between the dashed and solid Hilbert envelopes illustrates the effect of the filter on the antenna impulse response.

eled antenna function, found in Nelles et al. (2015). From this figure it can be seen that the filter slightly increases the rise time, produces some small oscillations just before the initial rise, and reduces the amplitude of the response function.

After filtering, any remaining RFI between 30 and 80 MHz is removed using the process described in Corstanje et al. (2016). RFI lines are found by looking at the relative phase information of each frequency bin. If a frequency bin is dominated by human RFI, then the relative phase between two close antennas is constant in time. This is in contrast to frequency bins that are not dominated by human RFI, where the relative phase is random and changes over time.



**Figure 6.** Top panel shows the frequency content of one block of data before filtering and radio frequency interference (RFI) cleaning. The bottom panel shows the same data after filtering and RFI cleaning. The dashed line shows the shape of the filter used on the data (scaled so that it is visible). Note that this figure is less clean than Figure 4, this is because the data shown in this figure are only for a single data block in order to detail precisely the result of RFI filtering.

multiple peaks. In this work we assume that each peak is due to a single point source. In the case of multiple peaks in a pulse, such as those in Figure 8, we only use the time of the strongest peak.

#### 4.1. RFI Mitigation and Pulse Finding

During the initial processing of the LOFAR data, we perform RFI mitigation and pulse finding. This process is applied to the time series data recorded by every active LBA. For each antenna, we process the data in blocks that are typically 65536 ( $2^{16}$  samples, also 327  $\mu$ s) time samples long. In order to avoid edge effects from the RFI mitigation, each block overlaps with the previous block by 50% and we keep only the pulses that are found from the middle 50% of each block.

##### 4.1.1. RFI Mitigation

During RFI mitigation, the initial and final 10% of the block is multiplied by a half-Hann window, and the data block is Fourier transformed and band-pass filtered between 30 and 80 MHz, since most of the human generated RFI is below 30 MHz and above 80 MHz in the LOFAR data. The filter is a block filter that has been convolved with a Gaussian with a width of  $\sigma_{\text{tapering}} = 2.5$  MHz in order to reduce artificial oscillations in the processed data. Because of the convolution, the filter has a value of 0.5 at 30 and 80 MHz. Figure 5 shows the effect of this filter on the impulse response of the antenna. The impulse response was calculated by taking the inverse Fourier transform of the mod-

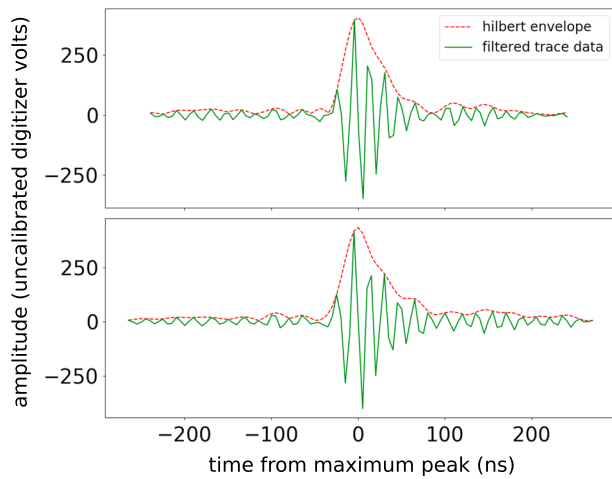
Therefore, we identify human RFI by looking at the stability of relative phases over a number of data blocks during time where there is no lightning. We can calculate the relative phase for each antenna and each data block, given by

$$\phi_{j,k}(\omega) = \arg(x_{j,k}(\omega)) - \arg(x_{j=0,k}(\omega)) \quad (1)$$

and the phase variance,

$$s_j(\omega) = 1 - \frac{1}{N} \left| \sum_{k=0}^{N-1} \exp(i\phi_{j,k}(\omega)) \right| \quad (2)$$

where  $\phi_{j,k}(\omega)$  is the relative phase for the  $j$ th antenna and  $k$ th block as a function of frequency,  $\omega$ .  $x_{j,k}(\omega)$  is the complex valued Fast Fourier Transform of the  $j$ th antenna and  $k$ th block as a function of frequency. There are  $N$  blocks of data analyzed for each antenna (typically  $N = 20$ ), and the  $j = 0$  antenna is the reference antenna that the phase is relative too. There is one reference antenna chosen per station.  $s_j(\omega)$  is the phase stability of the  $j$ th antenna as a function of frequency.  $\arg(x)$  returns the phase of the argument.  $\exp(x)$  is just  $e^x$ , and  $i$  is the imaginary number.



**Figure 7.** Two pulses recorded on two different LOFAR antennas, 35 km apart. Both of these pulses were from the same event. This is a simple event that has produced only a single peak within each pulse.

first of the five data points is the start of the pulse. If this is not the case, then the previous five data points are checked until this condition is reached. A similar method is used to find the end of the pulse. Because of this algorithm, the length of the resulting pulse is always the same in both polarizations for each pulse. We have found, using this algorithm, that the lengths of the found pulses tend to be between 100 ns and over 2,000 ns long. Figures 7 and 8 each show a pulse on two antennas from the same event resulting from this process. The two antennas used in Figures 7 and 8 are about 35 km apart.

The time of a pulse is found separately for the two polarizations. For each polarization, the five points of the Hilbert envelope centered on the highest peak are fit to a parabola. If a pulse has multiple peaks (such as the two pulses shown in Figure 8), the time of pulse is always considered to be the time of the highest peak, and the smaller peaks are not used.

#### 4.2. Station Timing Calibration

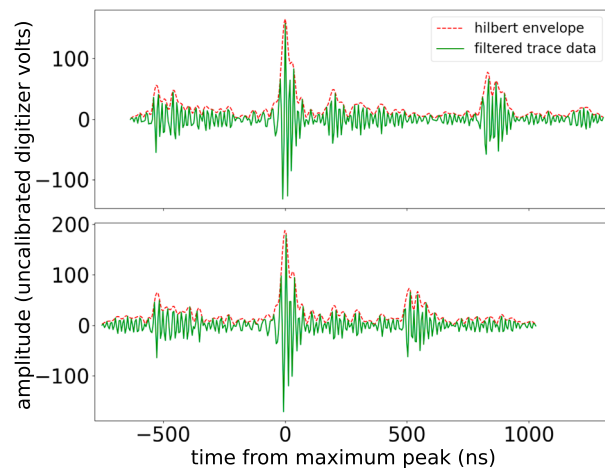
In the next step we find the timing offsets between LOFAR stations. This step is split between four substeps: finding single-station plane waves, grouping plane waves on the core stations into point source events,

correlating the plane waves in the remote stations to the known point source events, and then simultaneously fitting the location of all point sources and the station timing offsets. The processes in this step are outlined in Figure 9.

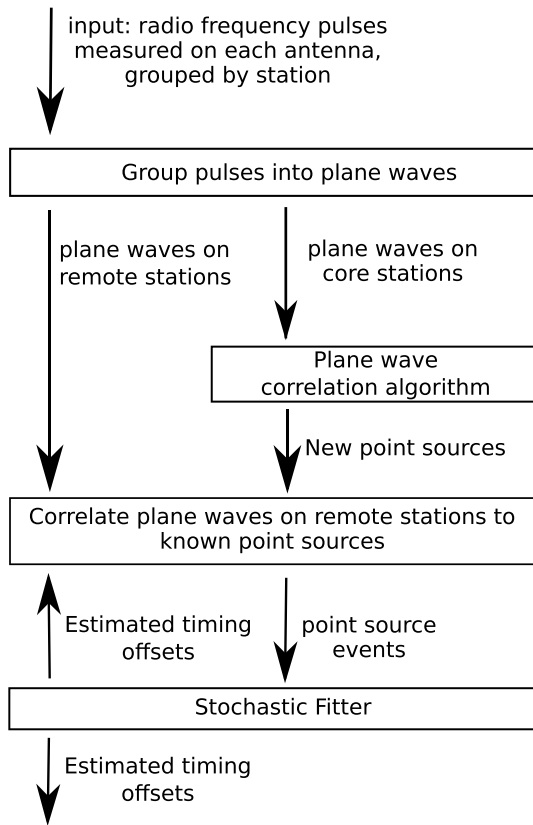
##### 4.2.1. Single-Station Plane Waves

Since each station is only about 60 m in diameter, an individual station only detects a plane wave from each lightning event. So in order to reduce the complexity of the data, all pulses recorded at each station are grouped into plane waves independently at each station. The algorithm that groups pulses into plane wave events has a search time that is initially set to the time of the first pulse observed at that station, on any polarization. Then the algorithm counts the number of antennas that have pulses in a bin that starts at the search time and has  $D/c$  duration in time, where  $D$  is the diameter of the station and  $c$  is the speed of light in air, accounting for the index of refraction of 1.000293 at standard temperature and pressure (Rüeger, 2002). If four or more antennas have pulses in this bin, then we say that we have found a plane wave event and we fit the times of the pulses on the different antennas to a plane wave model. This is done by minimizing  $F$ , which is the sum of squares of the difference between the model and the data,

$$F(t_p, \phi, \theta) = \sum [(x_c - x_i) \cos(\phi) \sin(\theta) + (y_c - y_i) \sin(\phi) \sin(\theta) + (z_c - z_i) \cos(\theta) - (t_i - t) \times c]^2 \quad (3)$$



**Figure 8.** Two pulses recorded on two different LOFAR antennas, 35 km apart. Both of these pulses were from the same event. Notice that because we are defining a pulse as the result of our pulse-finding algorithm, the entire time trace shown here is considered as “the pulse,” despite being a fairly complex event that has produced multiple peaks within each pulse.



**Figure 9.** The algorithm used to find the timing offsets between LOFAR stations. The blocks represent the pieces of the algorithm and the arrows represent the data passed between them.

with a nonlinear Levenberg-Marquard minimizer, where  $x_c$ ,  $y_c$ , and  $z_c$  is the center of the station.  $t_i$  is the time of the pulse received by the  $i$ th antenna.  $x_i$ ,  $y_i$ , and  $z_i$  is the position of the  $i$ th antenna.  $t_p$ ,  $\phi$ , and  $\theta$  are the arrival time and arrival direction (azimuth and zenith) of the plane wave and are the fitted parameters.  $c$  is the speed of light in air. Regardless if a plane wave is found or not, the search time is then advanced to the next pulse not associated with a found event and the algorithm repeats. For the July 2016 lightning flash we can find between 6,000 and 10,000 plane waves in each station.

Note that accounting for the index of refraction of air is critical, as not including it will produce timing errors of almost 1 ns per 1 km of propagation distance ( $\Delta t = R\Delta n/c_v$ , where  $\Delta t$  is the error in time,  $R$  is propagation distance,  $\Delta n$  is the error in index of refraction, and  $c_v$  is speed of light in a vacuum). Since LOFAR is capable of nanosecond timing precision, in the future we will need to consider the effects of temperature, pressure, and humidity. The most significant of these is pressure, since at 10 km altitude (pressure of 250 hPa) the index of refraction of air can be as low as 1.000064 (Rüeger, 2002).

#### 4.2.2. Grouping Core Stations

After grouping pulses into plane waves for each station, the plane waves across different stations need to be grouped together. This is made particularly difficult due to the unknown timing offsets between the stations. In order to overcome this difficulty we first group the plane waves recorded by just the core stations into point source events. This is done with our plane wave correlation algorithm.

The plane wave correlation algorithm starts with a reference station and loops over every plane wave found in that station. Since we have already found the time, azimuth and zenith angles of the plane wave, then the plane wave can be thought of as a ray that points from the point source, at an unknown distance, to the associated station. Therefore, distances backward along the ray can be used as initial guesses for source locations, according to

$$\begin{aligned} x(R) &= \cos(\phi) \sin(\theta)R + x_c \\ y(R) &= \sin(\phi) \sin(\theta)R + y_c \\ z(R) &= \cos(\theta)R + z_c \\ t(R) &= t_p - R/c \end{aligned} \quad (4)$$

where  $x$ ,  $y$ ,  $z$ , and  $t$  are the guess location and time of the point source event.  $t_p$ ,  $\phi$ ,  $\theta$  are the arrival time and arrival direction (azimuth and zenith) of the plane wave.  $x_c$ ,  $y_c$ , and  $z_c$  are the location of the center of the reference station.  $R$  is the distance from the station to the point source event, and  $c$  is the speed of light in air.

If we have a guess for  $R$ , then equation (4) gives us a guess for the time and location of an event. Using this guess time and location, we can then calculate the theoretical time a pulse from that event should be received on every other antenna using

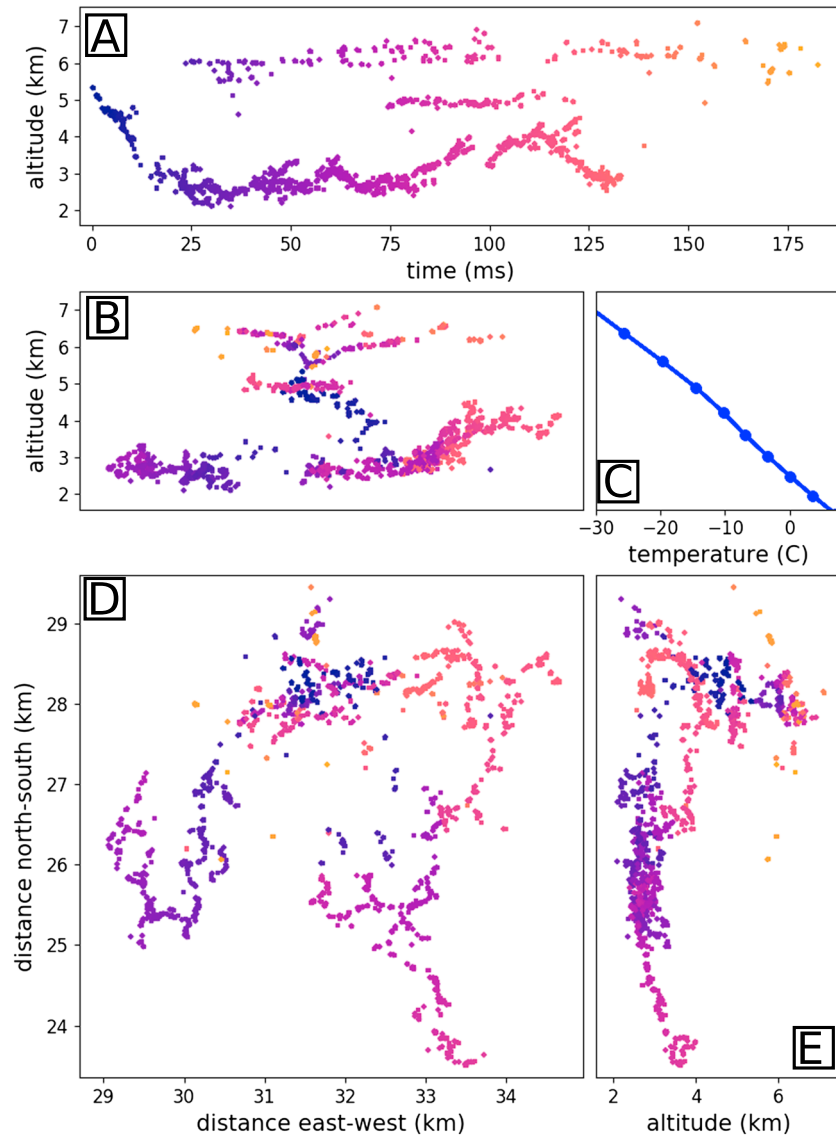
$$MT_i = t + \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2}/c \quad (5)$$

where  $MT_i$  is the model time that a pulse should be received by the  $i$ th antenna,  $x$ ,  $y$ ,  $z$ , and  $t$  is the location and time estimate given by equation (4), and  $x_i$ ,  $y_i$ , and  $z_i$  are the  $x$ ,  $y$ , and  $z$  coordinates of the  $i$ th antenna.

We can then estimate  $R$  by minimizing the root-mean-square (RMS) given by

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N (MT_i(x, y, z, t) - t_i)^2} \quad (6)$$



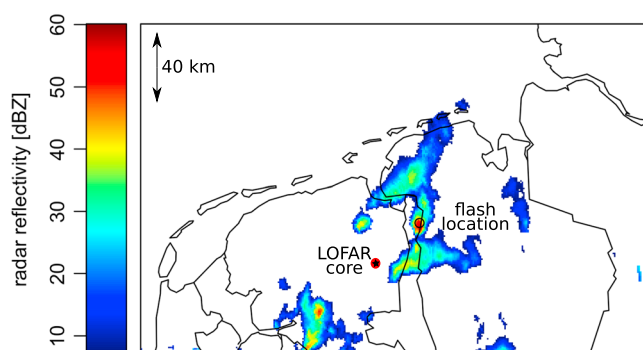


**Figure 10.** Result of our lightning mapping algorithm, applied to the July 2016 lightning flashes. Each point represents the location of an event. The locations of the sources found by the NW-SE polarizations, and the NE-SW polarizations are both shown. Panel a shows the altitude versus time. Panel b shows altitude versus east-west. Panel c shows altitude versus temperature, derived from Global Data Assimilation System data. Panel D is a plan view, showing distance east-west versus distances north-south. Panel E shows altitude versus north-south. The points are colored by time. The origin (0 km, 0 km, and 0 km) is in the core of LOFAR, the middle of station CS002. The color of the dots represents time.

with a nonlinear Levenberg-Marquard minimizer, where  $t_i$  is the time of the measured pulse that is closest to the model time. Note that we use RMS, instead of chi-square, because we do not have a good estimate on what the timing error should be. This is discussed more in section 6.

If the RMS fit value is below a certain threshold (typically 40  $\mu$ s), then we have found a point source event. If this is the case, we then polish the location of the point source event by fitting  $x$ ,  $y$ ,  $z$ , and  $t$  with a nonlinear minimizer. In order to avoid using a square root in a minimizer, we minimize  $F$  which, here, is the sum of the squares of the differences of the square of the propagation distance based on antenna location and time of the measured pulse,

$$F(x, y, z, t) = \sum [(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2 - c^2(t_i - t)^2]^2 \quad (7)$$



**Figure 11.** Composite of radar reflectivity from radar stations in De Bilt and Den Helder. This slice was taken at the lowest available altitude, which was 3,045 m for the De Bilt radar and 2,652 m for the Den Helder radar. The circle shows the area of the lightning flash.

The resulting location generally has an ill-determined radius (off by tens of km, with RMS fit values over 50 ns) because we still have not accounted for the timing offsets between the stations, and the entire LOFAR core lies within a circle of only 2 km radius.

Note that if the original direction of the reference plane wave is not correct, then this algorithm will not be able to find the associated pulses on the other stations that are due to the same event. Because of this, this algorithm is not very efficient at finding point sources. However, this is not a problem because at this point we are only searching for a small number of high-power events in order to calibrate the clock offsets between the different LOFAR stations.

#### 4.2.3. Correlating Remote Stations

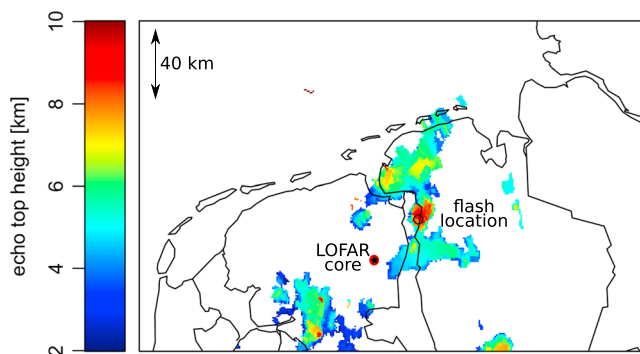
After grouping plane waves detected by the LOFAR core into point source events, we next attempt to correlate the plane waves detected by remote stations to the point sources already found in the previous step. At present, this step is mostly done by eye by plotting the times of plane waves mea-

sured by each station and the time that pulses should be seen from selected point sources. To make the match easier, we only consider the strongest plane waves and the strongest pulses, strong enough that the timing between the considered pulses tends to be tens of microseconds apart. These times between pulses, however, varies randomly. Therefore, the matches between the plane waves and the pulses tend to be unique.

If it proves too difficult to match the pulses and plane waves by eye, then this step can also be done by brute force. In which, for any given remote station, two point sources are selected from all known point sources. For each point source a plane wave is selected. For each point source/plane wave pair a station timing offset can be guessed by fitting the location of the source using the new plane wave and allowing the station timing delay to be a free variable. If the station timing delay calculated with both point sources is roughly the same, then we can say those two plane waves are due to the two associated point sources. Because of the large number of combinations of point sources and plane waves, this algorithm is extremely slow and is used only if absolutely necessary. Note that for the lightning flash mapped in this work, it was not necessary to use this brute-force method.

#### 4.2.4. Stochastic Station Timing Minimizer

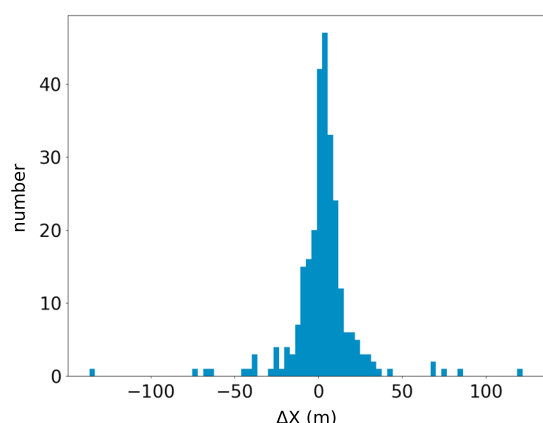
After finding as many point sources as possible across as many stations as possible, the station timing offsets can be found by simultaneously fitting the location and time of every point source and the timing offsets of every station using the model in equation (5) but with  $t_i$  now modified by the timing offset for the station that contains the  $i$ th antenna. In this model, however, the nonlinear minimizer has a tendency to get caught in a local minimum. To solve this, we run the minimization for a large number of runs (generally about 2,000), and for each run we perturb the initial guess by a random number drawn from a normal distribution. For the positions of the sources the normal distribution generally has a standard deviation of 100 m, and the standard deviation for the station timing guesses is generally 100 m/c.



**Figure 12.** Composite of cloud echo tops from radar stations in De Bilt and Den Helder. The echo tops are defined as the altitude where the radar reflectivity is 7 dB. The circle shows the area of the lightning flash.

In order to gain the most accurate estimate of the station timing offsets we make two cuts to the point sources that are correlated across core stations and remote stations. We only use point sources that have RMS fits less than 2.0 ns on every station individually, and we check the time traces of all the pulses that are grouped with each point source by eye in order to help insure that the correct pulses are grouped together. The time series of different events are sufficiently distinct that it is possible to distinguish between pulses that come from different events. Finally, we also require that every station contains pulses from at least 50% of the fitted point sources.

Since this stochastic minimizer drastically improves the estimated location of the point sources, reducing the radial error from tens of kilometers to meters, it also allows for more plane waves recorded by different stations to be associated with the known point sources in the general area of the lightning discharge. These new plane waves can then be

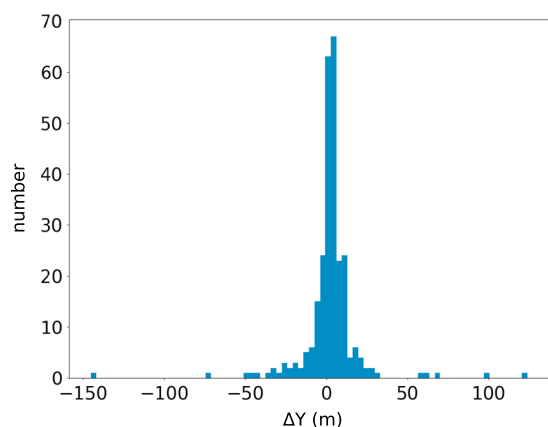


**Figure 13.** Histogram of difference in X (east/west) position between the two polarizations for 270 point sources that have root-mean-square fit better than 2 ns for both polarizations.

have a reasonable fit. The present algorithm we use to locate point source events we call the roaming hypersphere algorithm.

The roaming hypersphere algorithm is based on the idea that we can imagine a hypersphere that has a location  $\vec{X}$ , time  $T$ , and a radius  $R$  (units of seconds). We also have a number of antennas, where the location of the  $i$ th antenna is at  $\vec{X}_i$ . If there is a real point source at the center of the hypersphere, then it will produce a pulse on the  $i$ th antenna at time  $t_i = T + |\vec{X} - \vec{X}_i|/c$ . If there is an event inside the hypersphere, but not necessarily at the center, then that event will produce a pulse within the time bin of  $t_i \pm R$  on the  $i$ th antenna. Such a pulse we will say is “inside the hypersphere time bin.” In order to find the best guess for the location of a point source, we want to find the location of a hypersphere that has the smallest  $R$  but the largest number of antennas that have pulses inside the hypersphere time bin.

In order to utilize this technique, the hypersphere has an initial location that is our best guess for the location of the lightning flash. The initial radius of the hypersphere is very large, 20 km/c. Because this initial radius is so large, the initial flash location does not need to be very accurate, within 5 km easily suffices, and so we generally use the results from the previous step. The initial time is picked to be multiple  $R$  before the start of the lightning flash. The time of the hypersphere is increased by intervals of  $R/2$  until we find a large ( $>50$ ) number of antennas with pulses inside the hypersphere bin. Once such a time is found, we begin iterating to improve the guess of the event location. In each iteration we reduce the radius of the hypersphere so that only one pulse that was inside the hypersphere bin is now excluded from the hypersphere bin. Then we use a nonlinear minimizer to find a hypersphere position (radius held constant) that maximizes the number of antennas inside of the hypersphere bin. This loop continues until the radius of the hypersphere is 1,000 ns. If there are still a large number (50) of antennas that have pulses inside of the hypersphere bin, then we have found a point source.



**Figure 14.** Histogram of difference in Y (north/south) position between the two polarizations for 270 point sources that have root-mean-square fit better than 2 ns for both polarizations.

used to get a better estimate of the event locations and station timing offsets. In this way we repeat correlating plane waves to point sources and running the stochastic fitter until the best estimate on the station timing offsets can be found. About 10 point sources are sufficient to constrain the station timing offsets, but for the 2016 lightning flash we fit 64 point sources in order to get the best estimate of the station delays as possible. The fit does not significantly improve between fitting 32 and 64 point sources.

### 4.3. TOA Lightning Mapping

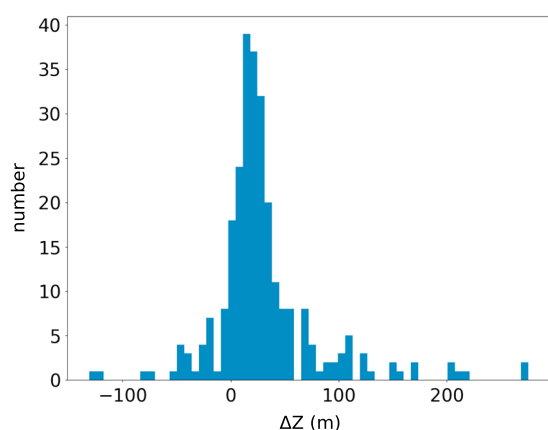
After the complex task of finding the station timing offsets, we attempt to locate as many point source events as possible. This is made difficult by the sheer amount of data collected by the LOFAR telescope. Each antenna can record over 10,000 pulses in a single flash, and we utilize about 150 antennas from the Dutch LOFAR stations. Because of this, and unlike more traditional LMAs, we cannot try fitting random combinations of pulses from different antennas and just keep the combinations that

have a reasonable fit. The present algorithm we use to locate point source events we call the roaming hypersphere algorithm.

The roaming hypersphere algorithm is based on the idea that we can imagine a hypersphere that has a location  $\vec{X}$ , time  $T$ , and a radius  $R$  (units of seconds). We also have a number of antennas, where the location of the  $i$ th antenna is at  $\vec{X}_i$ . If there is a real point source at the center of the hypersphere, then it will produce a pulse on the  $i$ th antenna at time  $t_i = T + |\vec{X} - \vec{X}_i|/c$ . If there is an event inside the hypersphere, but not necessarily at the center, then that event will produce a pulse within the time bin of  $t_i \pm R$  on the  $i$ th antenna. Such a pulse we will say is “inside the hypersphere time bin.” In order to find the best guess for the location of a point source, we want to find the location of a hypersphere that has the smallest  $R$  but the largest number of antennas that have pulses inside the hypersphere time bin.

In order to utilize this technique, the hypersphere has an initial location that is our best guess for the location of the lightning flash. The initial radius of the hypersphere is very large, 20 km/c. Because this initial radius is so large, the initial flash location does not need to be very accurate, within 5 km easily suffices, and so we generally use the results from the previous step. The initial time is picked to be multiple  $R$  before the start of the lightning flash. The time of the hypersphere is increased by intervals of  $R/2$  until we find a large ( $>50$ ) number of antennas with pulses inside the hypersphere bin. Once such a time is found, we begin iterating to improve the guess of the event location. In each iteration we reduce the radius of the hypersphere so that only one pulse that was inside the hypersphere bin is now excluded from the hypersphere bin. Then we use a nonlinear minimizer to find a hypersphere position (radius held constant) that maximizes the number of antennas inside of the hypersphere bin. This loop continues until the radius of the hypersphere is 1,000 ns. If there are still a large number (50) of antennas that have pulses inside of the hypersphere bin, then we have found a point source.

Once we find a point source we can find its location in a more traditional manner by fitting the times of the pulses on each antenna to the point source model in equation (5). If there are multiple pulses recorded by an antenna (inside of the hypersphere bin with a radius of 1,000 ns), then the pulse with the best fit to the point source model is used. Since this location fit is much more accurate than the center of the hypersphere, often more antennas with pulses within 1,000 ns of the model time can be found. Therefore, we iterate over fitting the location of the point source and looking for pulses within 1,000 ns on each antenna until the number of antennas included in the fit becomes stable. For simplicity, we use the times of the pulses on only the NW-SE polarized dipoles to find the point source events in the above algorithm. Once a point source is located in the data, two separate locations are fit to the times of the pulses in the NW-SE dipoles and the NE-SW dipoles independently, and the associated pulses are removed from the roaming hypersphere algorithm.



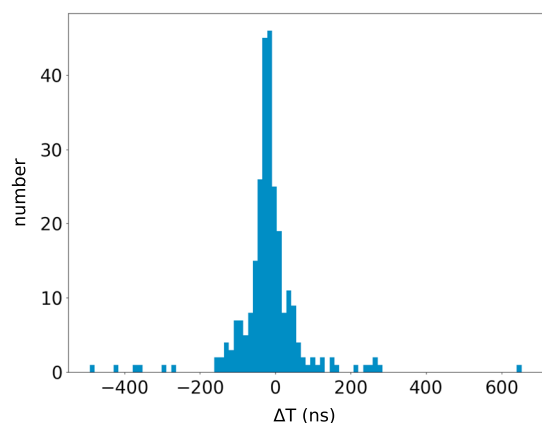
**Figure 15.** Histogram of difference in altitude between the two polarizations for 270 point sources that have root-mean-square fit better than 2 ns for both polarizations.

Unfortunately, this algorithm is not very efficient at locating point sources. Despite detecting 6,000–10,000 plane waves in each station, the roaming hypersphere algorithm currently finds 1,500 point sources for the July 2016 flash. Finally, since the first attempt at finding the station timing delays generally yields very few point sources, and only at great effort, the best fitting point sources from this last step can be fed back into the stochastic fitter, described in section 4.2.4, to refine the estimated station timing offsets. These refined station timing offsets can then be used to generate a better (more sources with smaller fit value) map of the lightning flash.

**5. Lightning Mapping Results and Atmospheric Conditions**

Figure 10 shows the results of our analysis applied to the lightning flash that was detected on the 12 of July 2016, at 17:34:55.100 UTC. This is plotted in a layout similar to that used for LMA data (e.g., Rison et al., 1999). The data have been filtered to show only events with RMS fits less than 5 ns. Figure 10 shows about 680 point sources located by the NW-SE polarized dipoles and 670 point sources located by the NE-SW polarized dipoles. Note that the Netherlands consists mostly of flat farmland and that the altitude of  $Z = 0$  m is a good approximation for the location of the ground, which is approximately at sea level. Figure 10 shows a number of common lightning processes. In the altitude versus time plot we can see an initial leader that starts around time 0 s and propagates downward until about 20 ms, after which it begins to propagate horizontally between 2 and 4 km in altitude until 140 ms. Figure 10 also shows the opposite end of the leader propagating horizontally between 6 and 7 km altitude from time 25 ms until after 175 ms. From these two main areas of horizontal propagation, we can surmise that there was a region of charge between 2 and 4 km altitude and a region of opposite charge between 6 and 7 km altitude. There is also a third region of horizontal leader propagation at about an altitude of 5 km between times 75 ms to about 120 ms. Since the initial downward leader propagated through the 5 km altitude region, then the horizontal propagation at 5 km must have branched off of the initial downward leader. Therefore, there must be a middle region of charge at 5 km altitude that is the same sign as the lower (2–4 km altitude) region of charge. We also notice from Figure 10 that the initial downward leader, and the subsequent horizontal propagation between 2 and 4 km, has a higher RF source density than the horizontal leader propagation between 6 and 7 km altitude.

Note that, so far, we have assumed that each of the located events is due to leader propagation. However, it is likely that some of these events are due to recoil leaders, as we have not located enough sources to distinguish between recoil leaders and normal leader propagation. Even if some of the events are due to recoil leaders, it should not invalidate our conclusions about locations and types of leaders and charge regions.



**Figure 16.** Histogram of difference in time between the two polarizations for 270 point sources that have root-mean-square fit better than 2 ns for both polarizations.

Based on the fact that the initial leader of intracloud lightning flashes tends to be negatively charged and that negatively charged leaders tend to produce a higher density of sources than positive leaders when detected by LMAs (Behnke et al., 2014), it is most likely that the initial downward leader was negative, so then the lower and middle charge regions were positive and the upper charge region, between 6 and 7 km altitude, was negatively charged. Figure 11 shows the radar reflectivity, and Figure 12 shows the cloud echo top height over the location of the lightning flash. The maximum cloud top height measured by SEVIRI (Spinning Enhanced Visible and Infrared Imager, see Roebeling et al., 2006) was 9.562 km. Based on the cloud top height, and the tendency of thunderstorms to have a tripolar structure, it is most likely that there was a positive charge region above 7 km that the flash did not propagate into. Figure 10 shows a temperature profile extracted from the Global Data Assimilation System (see <https://ready.arl.noaa.gov/gdas1.php>). This profile shows that the  $0^{\circ}$  C isotherm was at an altitude of about 2.5 km, the  $-10^{\circ}$  C isotherm was at an altitude of about 4.2 km, and the  $-20^{\circ}$  C isotherm was at 5.6 km altitude. Comparing these altitudes to the flash structure shows that the lower charge region was at the  $0^{\circ}$  C isotherm and the upper charge region was above the  $-20^{\circ}$  C isotherm.



**Table 1**  
*Three-Dimensional Leader Speeds*

Number	Inferred leader polarity	Location	Speed (m/s)	Number events
1	-	Initial downward leader	$1.51 \times 10^5$	37
2	-	Middle layer	$0.80 \times 10^5$	13
3	-	Middle layer	$0.64 \times 10^5$	10
4	-	Mower layer	$0.86 \times 10^5$	16
5	-	Lower layer	$0.81 \times 10^5$	7
6	+	Upper layer	$0.19 \times 10^5$	17
7	+	Upper layer	$0.11 \times 10^5$	7

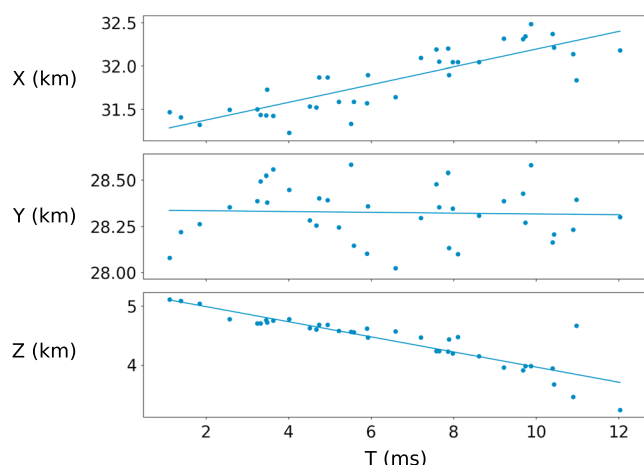
This is slightly different than Pilkey et al. (2014), who showed that (for Florida thunderstorms) the lower positive charge region tends to be around the  $0^\circ$  C isotherm, but the negative charge region tends to be between the  $0^\circ$  C isotherm and the  $-10^\circ$  C isotherm.

Since each point source has a separate fit to the NW-SE polarized dipoles and the NE-SW polarized dipoles, we analyzed the difference between the locations of the two polarizations for all events where both polarizations had a RMS fit better than 2 ns. Figures 13–16 show histograms of the position differences between the two polarizations. We found that the standard deviation of difference between the locations of the two polarizations was 20 m in X, 18 m in Y, 49 m in Z, and 93 ns in T. The average of the location differences was practically zero for the X, Y, and T coordinates, but the differences in the Z coordinate have an average of 30 m. This will be discussed further in section 7.

We have also analyzed the propagation speed of seven sections of the leaders, including the initial downward leader and two sections from each of the three layers. This was done by finding small sections of the leader that propagated in a nearly linear fashion and fitting the X, Y, and Z positions versus time with a line. The slopes of these lines were then used to estimate the 3-D propagation speed of the leader. We used the positions of only the events that were located by the NW-SE oriented dipoles and had timing errors smaller than 2 ns. Table 1 shows the analyzed leader speeds, along with the inferred charge of the leader, location of the leader segment, and the number of sources that were used to find the leader speed. The initial downward negative leader had the highest speed,  $1.51 \times 10^5$  m/s, and the horizontal negative leaders propagated at about half that of the initial downward leader. These negative leaders speeds are similar to previous measurements (on the order  $1.0 \times 10^5$  m/s (Dwyer & Uman, 2014)). The two positive leader segments were an order-of-magnitude slower than the negative leaders. This is unusual, as most positive leaders tend to propagate at the same speed as the negative leaders (Dwyer & Uman, 2014). However, a few upward positive leaders have been observed to propagate this slowly (Dwyer & Uman, 2014). van der Velde and Montanyá (2013), in particular, measured the speed of a number of lightning leaders using a LMA. They found that negative leaders tend to propagate around  $10^5$  m/s, and horizontal positive leaders tend to propagate around  $2 \times 10^4$  m/s, which is consistent

**Table 2**  
*Leader Velocities and  $R^2$  Values*

Number	X vel. (m/s)	Y vel. (m/s)	Z vel. (m/s)	X $R^2$	Y $R^2$	Z $R^2$
1	$0.93 \times 10^5$	$-0.10 \times 10^5$	$-1.18 \times 10^5$	0.66	0.04	0.68
2	$0.63 \times 10^5$	$-0.48 \times 10^5$	$0.03 \times 10^5$	0.93	0.91	0.03
3	$-0.60 \times 10^5$	$-0.17 \times 10^5$	$-0.14 \times 10^5$	0.95	0.66	0.54
4	$-0.79 \times 10^5$	$-0.30 \times 10^5$	$0.12 \times 10^5$	0.99	0.82	0.31
5	$-0.79 \times 10^5$	$-0.16 \times 10^5$	$0.05 \times 10^5$	0.99	0.82	0.06
6	$0.16 \times 10^5$	$0.04 \times 10^5$	$0.09 \times 10^5$	0.86	0.71	0.84
7	$-0.05 \times 10^5$	$-0.10 \times 10^5$	$-0.01 \times 10^5$	0.97	0.90	0.06



**Figure 17.** X, Y, and Z versus time of the sources used in segment 1, the initial downward negative leader, to estimate the leader speed. The linear fits are also shown.

with our measurements. Table 2 shows the velocities, in all 3 dimensions, of the leader segments, as well as the  $R^2$  coefficient of determination in each dimension, defined as follows:

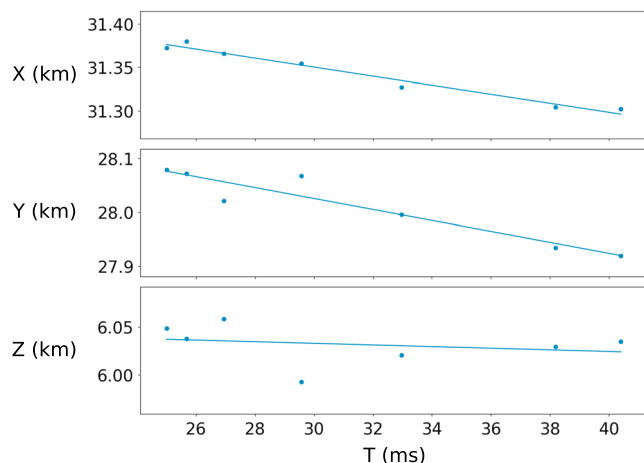
$$R^2 = 1 - \frac{\sum (y_i - f_i)^2}{\sum (y_i - \bar{y})^2} \quad (8)$$

where  $y_i$  is the  $i$ th data point,  $f_i$  is the value of the fitted line at the  $i$ th data point, and  $\bar{y}$  is the average of the data. If the  $R^2$  value is close to 1, then the leader position in that dimension was very linear with respect to time. If the  $R^2$  value is close to 0, then the leader position was not well fit by a line. Note that it is not problematic, it is even expected, if the  $R^2$  value is small when the leader velocity in that direction is small compared to the total speed. Figures 17 and 18 show the X, Y, and Z versus time for the sources and linear fits used in estimating the leader speeds for segments 1 and 7. Finally, it is possible that these leader speed measurements could have been effected by fast processes, such as recoil leaders, for which we do not have the source density to resolve. It is not clear how recoil leaders (if they were present) effect our analysis of the leader speed.

## 6. Simple Error Analysis

The largest source of timing error is due to the fact that the shape of the pulses measured by LOFAR changes between stations, as opposed to instrumental noise. Two primary manifestations of this effect are the electric field changing polarization and different peaks changing their relative timing within a single pulse. At present this is not well understood, and so we do not have a good way to estimate our timing error. We believe that the pulse shapes changing between antennas is due to the fact that LOFAR is sensitive to the spatial extent of the source region, and a point source model is no longer a good approximation. For this same reason it is not useful to use cross-correlation techniques as employed in lightning interferometers, which have small enough baselines that this is not a problem. In lieu of a thorough error analysis, we have performed a simple error analysis using a Monte Carlo technique. In this analysis we have used 64 point sources, the same 64 point sources originally used in the stochastic fitter to find the station timing offsets. We performed 1,000 runs, where we introduced an additional random timing error to the time of each measured pulse. This additional error was drawn separately for each antenna for each point source event from a normal distribution with a standard deviation of 2 ns. We used 2 ns since we originally selected point sources that have RMS fits less than 2 ns on every LOFAR station; therefore, the timing error of each pulse must be no more than 2 ns. After adding

in the additional error, we then fit the location and time of every point source and the timing offset of every station simultaneously.



**Figure 18.** X, Y, and Z versus time of the sources used in segment 7, in the upper layer, to estimate the leader speed. The linear fits are also shown.

From these fits we extract three different kinds of errors. First, we extracted the error in the station timing offsets by simply taking the standard deviation of the fitted timing offset for each station. These errors are shown in Table 3 for the core stations and Table 4 for the remote stations. These tables show that the station timing offsets have errors less than 1 ns for core stations and between 3 and 30 ns for remote stations, demonstrating that our technique significantly improves on the known cable timing delays. We also extracted the absolute position error for the whole flash. This was done by averaging the X, Y, Z, and T solutions over all 64 point sources for each run. The errors were just the standard deviations of the average X, Y, Z, and T, which are 10.3 m, 8.8 m, 67.9 m, and 30.0 ns, respectively. Finally, we found the relative location errors by taking standard deviation of X, Y, Z, and T for each point source after subtracting off the average X, Y, Z, and T for that run. Table 5 shows the average, standard deviation, minimum, and maximum of the errors of the 64 point sources. We also projected the errors of the X and Y coordinates into errors in the azimuthal direction and along the cylindrical radius. Table 5 shows our

**Table 3**  
Timing Offset Errors Using Lightning Data on Core Stations

Station	Timing offset error (ns)
CS001	0.23
CS004	0.15
CS006	0.18
CS011	0.26
CS013	0.25
CS021	0.38
CS026	0.59
CS028	0.51
CS030	0.59
CS031	0.53
CS032	0.44
CS302	0.96

**Table 4**  
Timing Offset Errors Using Lightning Data on Remote Stations

Station	Timing offset error (ns)
RS205	3.02
RS208	8.20
RS307	5.83
RS406	8.47
RS503	1.91
RS509	36.68
RS106	6.29
RS305	3.58
RS306	4.49
RS407	12.54
RS508	29.61

**Table 5**  
Location and Timing Errors

Coordinate	Average error	Standard deviation of error	Minimum	Maximum
X (m)	1.28	0.81	0.75	6.22
Y (m)	0.88	0.97	0.31	7.23
Z (m)	16.2	6.68	5.53	35.29
T (ns)	4.82	4.06	2.50	30.66
Azimuthal (m)	0.489	0.11	0.41	1.29
Horizontal radial (m)	1.4	1.27	0.70	9.47

**Table 6**  
Colorado LMA Location Accuracies at 5 km Altitude, From Chmielewski and Bruning (2016)

Distance from center (km)	Range precision (m)	Altitude precision (m)
5	2.53	18.50
10	2.63	20.51
20	2.82	30.91
30	3.53	40.91
40	5.23	40.22
50	7.80	38.56

worst error is in the vertical direction, which is normal for lightning mapping systems (Thomas et al., 2004). Future work will be needed to refine our error analysis and determine the best precision that can be achieved by LOFAR.

## 7. Discussion and Conclusion

We have developed a pipeline that can map lightning, similar to a LMA, using data collected by the LOFAR radio telescope. We have demonstrated that we can map lightning that is 40 km from the core of LOFAR, 30 km from the nearest station, and outside of the area enclosed by LOFAR with a relative precision on the order of 1 m in X and Y, 20 m in altitude, and 20 ns in time. Thomas et al. (2004) have shown that, for LMAs, the location accuracy improves drastically for sources located inside of the enclosed area, as opposed to outside the enclosed area as we have done in this study. Therefore, we can expect even better location accuracies for lightning flashes directly overhead of LOFAR. The Dutch LOFAR stations cover an area of about 3,200 km<sup>2</sup>, which is comparable to most LMAs. For comparison, the Colorado LMA is one of the largest LMAs in the world, and consists of 20 stations, and covers an area of 7,500 km<sup>2</sup> (*personal communication*, January 2018). Chmielewski and Bruning (2016) used a Monte Carlo simulation to estimate the location accuracy of the Colorado LMA. The horizontal range and altitude errors at a number of distances from the center of the Colorado LMA, at 5 km altitude, are shown in Table 6. Note that all of these locations are within the Colorado LMA's enclosed area. Thomas et al. (2004) used a balloon with a radio beacon to experimentally determine the New Mexico Tech LMA location accuracies for sources inside of the LMA's enclosed area. From the balloon data they found that the best location accuracy of the New Mexico Tech LMA is about 20 m in altitude and about 8 m in both horizontal coordinates.

Our present algorithm was able to locate seven sources per ms of data (1,500 located sources over 200 ms). This is already better than LMAs which can locate 2–3 sources per ms of data (5,000 sources over 2 s from Thomas et al. (2004), and 1,100 sources over 600 ms from Stock et al. (2014)). However, the station closest to the lightning detected over 10,000 pulses during the course of the flash. If we improve our algorithm to use all of these pulses in locating sources, we should be able to map around 50 sources per ms of data. This is comparable to lightning interferometers which can locate about 100 sources per ms of data (62,000 sources over 600 ms from Stock et al. (2014)). This should improve even further for lightning that is relatively closer to LOFAR.

It is interesting to note that we can achieve meter-level location accuracy with LOFAR, but we know that lightning leader steps have sizes on the order of tens of meters (Dwyer & Uman, 2014). Furthermore, the point sources are located independently on the two different polarizations, and the differences between the locations of the two polarization are significantly larger than our estimated location errors. From these two observations, we conclude that LOFAR is sensitive to the spatial extent of the current distribution of the individual leader steps. Therefore, since we also save the full trace data from a large number of dual polarized antennas that are distributed over a large area, LOFAR data can be used to probe the internal structure of lightning leader propagation for the first time.

We did find that one polarization tends to have systematically lower altitudes than the other. It is not at all clear why this is the case. We know that the NW-SE oriented dipole antennas are more sensitive to the Z component of the electric field than the NE-SW dipoles (for sources coming from the direction of the 2016 flash). It is possible that the horizontal component is damped out more over longer propagation distances due to the conductivity of the Earth, causing the two dipole orientations to receive the peak electric field at different times, which could then affect the reconstructed altitude of the event.

In future work we will develop techniques that can leverage LOFAR's full capability of probing the internal structure of leader propagation, potentially accounting for a finite conducting Earth and changing index of refraction of air. Undoubtedly, this will be a challenging task. After this is done, these techniques can hopefully be used to improve the TOA mapping we have introduced here, particularly to improve the efficiency that we can find point sources in the data. We also plan to investigate the potential of mapping lightning using interferometric techniques with LOFAR and to use the LOFAR HBAs to explore a higher-frequency regime. Finally, once the temporal and spatial structure of the lightning waveforms measured by LOFAR are understood, we will perform a thorough error analysis.



## Acknowledgments

The LOFAR cosmic ray key science project acknowledges funding from an Advanced Grant of the European Research Council (grant FP/2007-2013/ERC grant 227610). The project has also received funding from the ERC under the European Union's Horizon 2020 research and innovation program (grant 640130). We further acknowledge financial support from FOM (FOM grant 12PR304) and NWO (VENI grant 639-041-130). A. N. is supported by the DFG (grant NE 2031/1-1). This paper is based (in part) on data obtained with the International LOFAR Telescope (ILT) under project code LC6\_003. LOFAR (van Haarlem et al., 2012) is the Low Frequency Array designed and constructed by ASTRON. It has observing, data processing, and data storage facilities in several countries, that are owned by various parties (each with their own funding sources), and that are collectively operated by the ILT foundation under a joint scientific policy. The ILT resources have benefited from the following recent major funding sources: CNRS-INSU, Observatoire de Paris and Université d'Orléans, France; BMBF, MIWF-NRW, MPG, Germany; Science Foundation Ireland (SFI), Department of Business, Enterprise and Innovation (DBEI), Ireland; NWO, The Netherlands; The Science and Technology Facilities Council, UK. The data used in this paper are available from Hare et al. (2017).

Finally, we would like to thank Earle Williams for his extensive referee report.

## References

- Behnke, S. A., Thomas, R. J., Edens, H. E., Krehbiel, P. R., & Rison, W. (2014). The 2010 eruption of Eyjafjallajökull: Lightning and plume charge structure. *Journal of Geophysical Research: Atmospheres*, 119, 833–859. <https://doi.org/10.1002/2013JD020781>
- Chmielewski, V. C., & Bruning, E. C. (2016). Lightning mapping array flash detection performance with variable receiver thresholds. *Journal of Geophysical Research: Atmospheres*, 121, 8600–8614. <https://doi.org/10.1002/2016JD025159>
- Corstanje, A., Buitink, S., Enriquez, J. E., Falcke, H., Hörandel, J. R., & Krause, M. (2016). Timing calibration and spectral cleaning of LOFAR time series data. *Astronomy and Astrophysics*, 591, A41. <https://doi.org/10.1051/0004-6361/201527809>
- Dwyer, J. R., & Uman, M. A. (2014). The Physics of Lightning. *Physics Reports*, 534, 147–241. <https://doi.org/10.1016/j.physrep.2013.09.004>
- Edens, H. E., Eack, K. B., Eastvedt, E. M., Trueblood, J. J., Winn, W. P., Krehbiel, P. R., et al. (2012). VHF lightning mapping observations of a triggered lightning flash. *Geophysical Research Letters*, 39, L19807. <https://doi.org/10.1029/2012GL053666>
- Hare, B. M., Scholten, O., Bonardi, A., Corstanje, A., Ebert, U., Falcke, H., et al. (2017). D20160712T173455.100Z LOFAR lightning flash, hdl:10411/QBXVY1, DataverseNL Dataverse, V1.
- Nelles, A., Hörandel, J. R., Karskens, T., Krause, M., Buitink, S., Corstanje, A., et al. (2015). Calibrating the absolute amplitude scale for air showers measured at LOFAR. *JINST*, 10, P11005. <https://doi.org/10.1088/1748-0221/10/11/P11005>
- Pilkey, J. T., Uman, M. A., Hill, J. D., Ngim, T., Gamera, W. R., Jordan, D. M., et al. (2014). Rocket-triggered lightning propagation paths relative to preceding natural lightning activity and inferred cloud charge. *Journal of Geophysical Research: Atmospheres*, 119, 13,427–13,456. <https://doi.org/10.1002/2014JD022139>
- Rison, W., Krehbiel, P. R., Stock, M. G., Edens, H. E., Shao, X. M., Thomas, R. T., et al. (2016). Observations of narrow bipolar events reveal how lightning is initiated in thunderstorms. *Nature Communications*, 7, 10721. <https://doi.org/10.1038/ncomms10721>
- Rison, W., Thomas, R. J., Krehbiel, P. R., Hamlin, T., & Harlin, J. (1999). A GPS-based three-dimensional lightning mapping system: Initial observations in central New Mexico. *Geophysical Research Letters*, 26, 3573–3576. <https://doi.org/10.1029/1999GL010856>
- Roebeling, R. A., Feijt, A. J., & Stammes, P. (2006). Cloud property retrievals for climate monitoring: Implications of differences between Spinning Enhanced Visible and Infrared Imager (SEVIRI) on METEOSAT-8 and Advanced Very High Resolution Radiometer (AVHRR) on NOAA-17. *Journal of Geophysical Research*, 111, D20210. <https://doi.org/10.1029/2005JD006990>
- Rüeger, J. M. (2002). Refractive index formulae for radio waves. In *Proc. FIG XXII International Congress*. Washington, DC.
- Scholten, O., Buitink, S., Dina, R., Dorosti Hasankiadeh, Q., Frieswijk, W., Hendriks, F., et al. (2017). Lightning imaging with LOFAR. *EPJ web of conferences*, 135, 03003. <https://doi.org/10.1051/epjconf/201713503003>
- Stock, M. G., Akita, M., Krehbiel, P. R., Rison, W., Edens, H. E., Kawasaki, Z., & Stanley, M. A. (2014). Continuous broadband digital interferometry of lightning using a generalized cross-correlation algorithm. *Journal of Geophysical Research: Atmospheres*, 119, 3134–3165. <https://doi.org/10.1002/2013JD020217>
- Thomas, R. J., Krehbiel, P. R., Rison, W., Hunyady, S. J., Winn, W. P., Hamlin, T., & Harlin, J. (2004). Accuracy of the lightning mapping array. *Journal of Geophysical Research*, 109, D14207. <https://doi.org/10.1029/2004JD004549>
- van der Velde, O. A., & Montanyá, J. (2013). Asymmetries in bidirectional leader development of lightning flashes. *Journal of Geophysical Research: Atmospheres*, 118, 13,504–13,519. <https://doi.org/10.1002/2013JD020257>
- van Haarlem, M. P., Wise, M. W., Gunst, A. W., Heald, G., McKean, J. P., Hessels, J. W. T., et al. (2012). LOFAR: The LOw-Frequency ARray. *Astronomy and Astrophysics*, 556, A2. <https://doi.org/10.1051/0004-6361/201220873>